DOCUMENT RESUME

ED 144 951 TM 006 265

AUTHOR Churchman, David; Hoepfner, Ralph
TITLE Tailoring A Testing Program to the Needs of Varied Users.
PUB DATE [Apr 77]
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Elementary Secondary Education; *Evaluation Needs; Military Organizations; *Needs Assessment; School Personnel; Schools; Student Testing; Surveys; Testing Problems; *Testing Programs
IDENTIFIERS *Overseas Dependents School System

ABSTRACT

School testing programs in many cases have been limited to obtaining an IQ score and achievement scores in reading and mathematics for each student. Testing in the U.S. Department of Defense Overseas Dependents Schools followed this pattern and was under attack from many sides. Consequently, the testing program was suspended in 1971 to provide funds for an evaluation to determine the most appropriate type of testing program to meet the needs of the Overseas Dependents Schools. A four step needs assessment evaluation provided the necessary information. First, 211 areas for testing in the elementary and secondary schools were identified. Second, the relative importance to teachers and administrators of having information in each of the areas was determined. Third, problems associated with the old testing program and characteristics of the Overseas Dependents Schools that warranted consideration in developing a testing program were identified. Fourth, the information was analyzed to determine the most appropriate purposes for testing at each administrative level of the system, and the most appropriate tests, sampling and administrative procedures for a testing program to provide the required information. (Author/MV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED144951

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

David Churchman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

SUMMARY

Tailoring a Testing Program to the Needs of Varied Users

David Churchman, California State University
Ralph Hoepfner, Systems Development Corporation

School testing programs commonly have been limited to obtaining little more than an IQ and achievement scores in reading and mathematics for each student. Testing of this type is under attack from many sides: some feel the tests are not valid predictors of important skills; minority groups argue that the tests are culturally biased; some fear that testing has a negative effect on self-esteem; others complain that tests do not measure critical qualities such as honesty or ambition; some argue that testing in curricular areas such as art and science is equally important, some feel that testing is an invasion of individual privacy; and others claim that teachers rely too rigidly on test results. Many of these complaints are justified; some reflect misunderstandings.

Testing in the U.S. Department of Defense Overseas Dependent Schools System (ODS) followed the traditional pattern and was subject to all of the attacks. The schools of the ODS primarily serve the minor dependent children of Department of Defense personnel. Centrally administered from the Pentagon by the Directorate for Dependents Education, the system is divided into three areas: Pacific area schools administered by the Air Force, Atlantic area schools by the Navy, and European schools (including schools located in the Middle East and Africa) by the Army.

TM006 265

With an enrollment of almost 175,000 students, it is one of the largest American school systems. However, the relative isolation from American life, the extreme mobility of the students, the frequent turnover of staff, difficulties of communications within the system, requirements imposed by the laws of host countries, and the vagaries of international relations pose problems that are unique among American schools.

The problem raised by the Department of Defense in its evaluation of the ODS testing program required a decision as to the most important purposes that it should serve. Determination of the important purposes of a testing program called for a needs-assessment evaluation that would collect information and process it objectively and would compare present with desired practice for each of the purposes identified. The steps of the needs-assessment evaluation are explained below as they were applied to the problem of evaluating the need for test information in the ODS.

Determining the Purposes the Testing Program Should Serve

In an election, a write-in candidate usually has no real chance of winning. The more complete the list of names on the ballot, the more likely the results will reflect the opinions of the electorate. Similarly, if teachers and administrators rate a complete list of areas in which testing might be conducted, the results would reflect the importance of various types of test information. Therefore, the first task in conducting the needs assessment was to draw up a ballot that presented a complete list of affective, cognitive, and psychomotor areas in which testing could occur.

Representative materials, including course syllabi, lists of textbooks used, and descriptions of curricular offerings were collected from schools throughout the ODS to develop the required ballot list.

A major problem in conducting this type of "election" is establishing a suitable level of specificity for the testing purposes. If they are too specific, they become trivial for guiding a testing program, and would be so numerous that the ballot would be too long to complete. If the purposes were very general they would be perceived to include areas of varying importance and thus would be too ambiguous to rate. Analysis of materials from the ODS led to development of a 106 item ballot for elementary schools and a 105 item ballot for secondary schools. Three purposes from each ballot are presented in Table 1 by way of examples. One ballot for elementary and one for secondary school personnel called for them to rate on a five-point scale the importance of having test information about each area. Complete lists of purposes on the two ballots are presented in Churchman, Alkin, Hoepfner, and Bradley (1972).

- - - - - - - - - - - - - - -

Insert Table 1 about here

- - - - - - - - - - - - - - -

Determining the Importance of Areas in Which Testing Should Occur

The results of any election depend to some extent on who is franchised to vote. In collecting needs-assessment information, everyone who will need the information should be able to express an opinion. The ODS wished to determine needs for test information at several levels, including the classroom teacher, the school, the overseas area offices, and the office

of the Pentagon. It was necessary, therefore, to obtain information from representatives of each of these levels. Schools administered by each of the three services were sampled, treating elementary and secondary schools separately. In this manner, all army-administered elementary schools in the European area, for example, were assigned to a single sampling cell, and a random sample of the schools in that cell was drawn. Similarly, each school in the ODS was assigned to its respective cell, and a random sample of schools in each cell was drawn.

These procedures ensured that schools throughout the ODS would be sampled, and that each school had an equal opportunity (within each cell) of being selected. Other dimensions of the sampling plan were the isolation and the size of the school. Isolation of the schools, as measured by distance from other dependent schools, time required to reach the school, number of visits from area-level personnel, hardship ratings, and the like, probably affect the attitude of the school personnel. School size is important because the number and type of specialized personnel and equipment at a school are largely determined by formulas that authorize "so much of this and so much of that" based on school enrollment.

Ballots were then sent to each school in the sample; 107 schools, or 89% of the sample, returned a total of 846 elementary and 677 secondary ballots. Means and ranks were computed as measures of the relative importance of each of the potential testing areas.

The most noteworthy finding was that there was little agreement among the schools in the samples. Thirty-five different curriculum areas appeared among the top five at one or more elementary schools; thirty-three different areas appeared among the top five at one or more secondary schools. The lack of agreement among schools in the samples as to the most important areas for which each needs test information suggests that it is not justified to select five or ten or some other arbitrary number of areas for testing throughout the entire ODS. Rather, each school should be free to determine some of the areas for which it desires test information.

Each potential area for testing listed in the ballot was conceived of as being part of a larger domain. The curriculum domains that appeared at the top and the bottom of the rankings are presented in Table 2.

- - - - - - - - - - - - - - -
Insert Table 2 about here
- - - - - - - - - - - - - - -

The unexpected characteristic of these results is the clear rejection of the importance of test information for foreign language skills, which contrasts sharply with the official concern for personal benefits from the overseas experience. The ballot findings suggest that tests related to areas such as honesty, creativity, persistence, critical thinking, sportsmanship, and similar individual attributes are not perceived as being important for testing.

Comparing Present with Desired Purposes for a Testing Program

From among the schools receiving the ballots, a smaller sample was selected for visitation. A three-part strategy for on-site interviews was developed that called for (1) identification of issues and concerns affecting testing programs at the first few schools visited, (2) in-depth discussion of one or two of these issues with each of the groups interviewed at the next few schools, and (3) verification of findings and interpretations at the last few schools visited. The third part of the strategy ensured that views expressed by individuals were common to more than one school and it allowed the interviewers to obtain reactions to specific suggestions made at other schools.

The interview schedule was completed by 375 ODS personnel. The first set of items in the schedule contained 19 criticisms of tests similar to those in the first paragraph of this article. The respondents indicated the extent to which they thought each criticism was a problem in testing children in their particular school. The second set contained 20 statements (19 for elementary personnel) that described various assessment and evaluation activities that occur in school settings. Respondents were to indicate how frequently they had used published tests for each activity. The third set of items repeated those of the second set, but respondents were asked how important it was for them to have that type of information in the future

While widespread disagreement was found among schools as to the curriculum areas for which test information was needed, there was strong agreement among those interviewed, regardless of grade level or geographic

area, as to the way tests were used and should be used. In particular, staffs felt that it takes too long after testing is completed to get scores back, and that the results depend too much on how students feel when they take the tests. Elementary personnel were much less satisfied with once-a-year testing than were secondary personnel.

Both groups felt that cultural bias was an important problem with the tests. Three distinct aspects of this problem were noted. First, there was the problem that members of minority groups have with the tests, which is the same as that in the United States. Second, there was the problem that children of an American soldier-father and a non-American (Korean, Vietnamese, German) mother had with the tests. Third, there were biases that stem from differences between civilian and military life. A test question that asks children to distinguish the picture of a store from those of a hospital and a school is difficult for children at military bases where buildings are sometimes architecturally indistinct. One item of one test asks the child to identify the way milk is delivered, and pictures a truck, a plane, and a boat. All three are correct, depending upon the base at which the child is stationed! While the number of such items is small, one or two such items on a subtest can have a significant effect upon the child's score.

Both the ballots and the interviews suggested that school staff did not view testing as an invasion of privacy, or a cause of excessive competition among students. However, they did view cheating as a major problem of testing, and students with whom this problem was discussed viewed competition, in addition to the repetitiveness of the tests and

the lack of information as to results, as a major cause of the cheating.

The interviewees discussed many other considerations they felt important in designing a successful testing program. The relative advantages of teacher vs. specialist administration of tests was raised, with the weight of opinion in the direction of teacher administration. The problems of once-a-year testing and multiple testing were weighed against student mobility patterns. The best hour of the day, length of test, and age at which separate answer sheets could be introduced, the type of training needed by teachers to improve administration of tests and factors such as the attitude of the instructor that would influence the success of the training were discussed. The ways that testing conditions varied across this worldwide system were explored in order to enable development of guidelines to improve overall uniformity and thus make test results more comparable. Problems of coordinating a testing program on a worldwide basis were discussed. The interviewers compared and interpreted their findings and developed a plan to account for as many of the problems as possible.

Determine Testing Procedures Most Important to Correct

It was apparent from the interviews that tests were not optimally used, nor did teachers want tests to be used for the purposes of grading students, promoting students, or accountability of teachers for student learning. Rather, there was a clear preference among teachers for test information that would provide diagnostic, placement, and counseling

information. It was unfortunate that the tests of the extant testing program were not designed for and would not be appropriate if used for those purposes.

There was little interest at the area and system level in using tests to evaluate individual teachers, but there was concern with evaluating priority and experimental programs, with identifying and clarifying problems at the ODS, and with reporting more completely the accomplishments of the system to the Congress of the United States.

The evidence accumulated from the ballot and the interviews suggested that four major discrepancies should be dealt with in designing the new testing program. First, few teachers used the information collected because it did not provide the diagnostic information they needed. Second, scores from the testing program provided no information about many important aspects of the curriculum such as art, music, commercial subjects, and the physical and life sciences, and thus fell far short of measuring the full range of achievement of the schools. Third, the testing program was inadequate for making decisions about the effectiveness of priority or innovative programs such as those dealing with minority groups, drugs, or career education. Fourth, even in those areas measured by the testing program, student mobility patterns in and out of the schools made it impossible to interpret the scores as measures of the schools themselves, because they did not identify the source of the learning that was measured.

Recommendations

Elimination or reduction of the discrepancies noted above requires that testing be conducted for six purposes.

A. At the school level, testing should be conducted to:

1. Diagnose students who appear to be having learning difficulties. This finding is especially crucial at the elementary level and for reading skills in all grades.
2. Place newly arrived students. The placement tests should be short, easily scored, and need yield only gross placement data.
3. Enable the school staff to evaluate their program at the local level.

B. At the district and area levels, testing should be conducted to:

4. Evaluate the implementation, progress, and impact of priority and experimental programs.
5. Clarify the nature of problems, such as those associated with the characteristics of troops assigned to a particular base.

C. At the Department of Defense level, testing should be conducted to:

6. Enable the Department of Defense to report to the Congress factual information about the effectiveness of the ODS.

In order to meet these requirements, specific recommendations were made with respect to teacher training, testing conditions, student attitudes, implementing the testing program, and the evaluation of program-evaluation systems within the school. Tests appropriate for elementary and secondary schools were identified, employing the CSE Test Evaluation series (Hoepfner,

et al, 1970, 1974), as being in the top priority curriculum areas of any school. A system to ensure that the self-evaluation was accountable to higher administration levels of the system was included.

The completed report was reviewed by a committee of Pupil Personnel Service directors and selected staff members of the ODS. Recommendations to allow schools more responsibility in developing their own testing programs, and to implement a sampling plan to meet ODS information needs were adapted to administrative requirements and a new program of testing more accurately tailored to the needs of the dependents education system has been implemented.[1]

---

[1]We would like to thank Dr. Thomas Drysdale, Deputy Director of Dependents Education, for reviewing this article and adding the note on the way the evaluation report was used by ODS.

Table 1

Sample Purposes of the Testing Program from the ODS Ballots

Elementary Level Testing Program	
GROUP ACTIVITY - SPORTSMANSHIP	Is a good winner and a good loser. Can be a leader or a follower. Obeys the rules of the game. Feels very involved in the sport. Has team spirit.
INFERENCE MAKING FROM READING SELECTIONS	Correctly interprets what is read. Recognizes from the material read what kind of characters are being talked about. Can tell that the characters in a story are sad, happy, trustworthy, or not to be trusted, etc. Can tell why characters act as they do.
SYSTEMATIC REASONING	Produces and solves complex problems and evaluates their solutions. Considers all the elements in situations and arrives at solutions through deductive reasoning.
Secondary Level Testing Program	
ATTITUDE TOWARD THE WORLD OF WORK	Is interested in understanding the world of work, both in its larger scope and as it relates to own chosen vocation.
WRITING	Expresses self clearly in writing, using adequate vocabulary and grammatically correct sentences. Knows the various types of writing (narrative, descriptive, argumentative, and persuasive), and organizes own writing. Knows and uses the rules governing various special written forms such as letters, applications, orders, and scientific reports.
VISUAL ARTS	Has a basic understanding of the nature and scope of the visual arts. Is familiar with the various media, techniques, and styles of the visual arts. Is familiar with historical and contemporary works of art in this and other cultures. Is able to analyze and criticize works of art.

Table 2

Most and Least Important Domains for the ODS Testing Program

Elementary Level Curriculum Domains	
Most Important	Least Important
Affective	Arts and Crafts
Language Construction	Foreign Language
Arithmetic Concepts	Music
	Religion

Secondary Level Curriculum Domains	
Most Important	Least Important
Mental Health	Mathematical Skills
Intellectual Functioning	Knowledge of Arts
Vocational Competence	Understanding of Nature
	Understanding of Technology

References

Churchman, D. A., Alkin, M.C., Hoepfner, R., & Bradley, P. A. An Evaluation of the Testing Program of the Dependents Overseas Schools of the Department of Defense (DAHC 15-73-C-0061). Los Angeles: Center for the Study of Evaluation, UCLA, 1974.

Hoepfner, R., Strickland, G., Stangel, G., Jansen, P., & Patalino, M. CSE Elementary School Test Evaluations. Los Angeles: Center for the Study of Evaluation, UCLA, 1970.

Hoepfner, R., Conniff, Jr., W. A., Hufano, L., Bastone, M., Ogilvie, V. N., Hunter, R., & Johnson, B. L. CSE Secondary School Test Evaluations: Grades 11 and 12. Los Angeles: Center for the Study of Evaluation, UCLA, 1974.

Hoepfner, R., Conniff, Jr., W. A., McGuire, T. C., Klibanoff, L. S., Stangel, G. F., Lee, H. B., & Rest, S. CSE Secondary School Test of Evaluations: Grades 9 and 10. Los Angeles: Center for the Study of Evaluation, UCLA, 1974.

Hoepfner, R., Conniff, Jr., W. A., Petrosko, J. M., Watkins, J., Erlich, O., Todaro, R. S., & Hoyt, M. F. CSE Secondary School Test Evaluations: Grades 7 and 8. Los Angeles: Center for the Study of Evaluation, UCLA, 1974.